下载APP

03 | 如何分析与处理音乐信号？

2021-11-26 冯建元

《搞定音频技术》 课程介绍 >

讲述：冯建元
时长 15:25 大小 14.13M

你好，我是建元。

上节课我们讲了语音信号分析的基础知识和方法，那么在常见的音频信号类型中，还有一种是音乐信号，这节课我们就一起来聊聊它。音乐信号受乐理、乐器发音规律、心理感知等因素的影响，与语音信号在分析和处理方法上不尽相同。而且音乐层面的分析涉及的知识领域比较多，知识容易发散。所以这节课我们会先介绍一些音乐信号的基本概念，然后从我们实时音频互动中，音乐场景遇到的案例出发，带你看看这些问题背后的原理和解决方法。

音乐信号中的基础概念

我们先来看看音乐信号中的一些基础概念。这里主要介绍一下常见乐器的发音原理，什么是速度和节拍，以及什么是音调和调式。

常见的乐器种类和发音原理

你还记得语音信号分析中，人的发音原理吗？音乐信号中除了人声歌唱的部分外，主要是乐器的声音，那么乐器是如何发音的，它们有什么规律吗？我们结合常见的乐器种类一起看下。

日常中常见的乐器主要有三种：打击乐器、弦乐器和管乐器。

第一种，打击乐器，比如我们常说的“敲锣打鼓”。**这类乐器的发音主要是依靠打击振动**。比如，鼓在鼓皮绷紧的状态下，受外力振动。振动的速度越快、力量越大、振幅越大，发出的声音就会越响亮。

打击乐器，比如鼓的音调，取决于鼓的尺寸、鼓皮的材质等。一个鼓经过调音，比如调节鼓皮的松紧程度后音调就固定了，所以如果需要同时演奏不同的音调就需要一组鼓，比如图 1 中的架子鼓。

图 1 图片来源于网络

第二种，弦乐器，包括钢琴、二胡、小提琴等乐器。**这类乐器发音来源于弦的振动**。不同形状的弦发出的声音特点也有所不同：若弦短而细，且绷紧，则乐器发出的音调高；若弦长而粗，且不紧，则乐器发出的音调低。

乐器的声音大小取决于弦的振幅大小。为了促使乐器能够发出洪亮的声音，一般会采用木制的共鸣箱以达到增强音量的效果。这一过程和人声浊音的发音十分相似，所以**弦乐器也会有基频和谐波**，**弦乐器的谐波也叫做泛音**。比如图 2 中的大提琴的声音，由于发音的基频和泛音的频率范围和人声相似，所以大提琴也被称为最像人声的乐器。

图 2 图片来源于网络

第三种，管乐器，包括笛子、号、萨克斯等乐器。**这类乐器发音来源于空气柱振动**。当吹奏管乐器时，抬起不同的手指会形成不同长度的空气柱，**若空气柱越长，则音调越低。若空气柱越短，则音调越高**。比如图 3 中的竖笛，越短的笛子发出的声音音调越高。

图 3 图片来源于网络

速度与节拍

知道了乐器的发音原理，接下来我们来看看乐曲是如何演奏的。

一首曲子的节奏是由速度和节拍决定的。速度我们一般以 BPM（beats per minutes）来表示。我们在演奏或者歌唱练习时，用的节拍器就是按照 BPM 来打节奏的。比如 80BPM，节拍器就会每分钟发出 80 次“嘀嗒”声，如果是用鼓来打节奏，就每分钟打出 80 个间隔均匀的鼓点。

而**节拍用来描述音乐中的进程的规律。**在音乐中，指有一定强弱分别的一系列拍子，在每隔一定时间重复出现。若干个这样有规律的拍子我们叫一个小节，比如常见的四拍一个小节。一首乐曲的节拍是作曲时就固定的，不会改变。一首乐曲可以是由若干种节拍相结合组成的。常见的节拍有四四拍 4/4、四二拍 2/4、四一拍 1/4、四三拍 3/4、八六拍 6/8 等。

大部分流行歌曲比如老鹰乐队的《Hotel California》、周杰伦的《东风破》都是 4/4 拍，而 3/4 拍的歌曲因为给人一种跳跃的感觉，所以多被用于舞曲，比如约翰·施特劳斯的《蓝色多瑙河》。在流行乐中也有一些慢舞曲，比如朴树的《白桦林》，也为四三拍。

所以，我们可以看到**乐器的演奏其实就是按照节拍规定的音符演奏顺序，然后按照指定的速度演奏出来**。

那乐曲中的音调和调式是如何从频率等数学的角度来分析呢？下面我们来看一下。

音调与调式

音调

基频和音调的英文都为 pitch，音乐信号中音调其实也是和乐器或者人声中的基频的频率一一对应的。比如，我们给乐器调音中常说的中央 C 就是基频频率约为 261.6Hz，唱名为 do，并且它位于乐音体系的最中央的位置，因而得名。中央 C 在国际标准中为 C4，在德国标准中为 c1，为了避免混淆，我们这里统一用国际标准来做解读。

那么我们平时说的 C 大调、D 大调和我们小时候音乐课的 do\re\mi\fa\so\la\ti 以及基频频率的关系是什么呢？

do\re\mi\fa\so\la\ti 是唱名，我们平时唱谱就是用这些音来把谱子唱出来的。他们和音名，也就是在音乐中包含的七个基本音调 CDEFGAB 一一对应。我们以钢琴中的中央音为例，唱名、音名和基频频率之间的对应如表 1 所示：

唱名	do	re	mi	fa	so	la	ti	do
音名	C4	D4	E4	F4	G4	A4	B4	C5
基频频率 (Hz)	261.6	293.7	329.6	349.2	392.0	440.0	493.9	523.2

极客时间

表1 中央C开始的大调唱名、音名和基频频率的关系

我们可以看到 C5 的基频频率正好是 C4 的两倍，这个其实就是我们说的度的概念，即 C5 比 C4 高八度，且一个八度其实就是基频频率相差一倍。高一个八度我们就把国际标准音名里后面的数字加 1。所以，比如 A 的音调从低到高可写为 A0、A1、A2 一直到 A9。

调式

我们知道了音调是什么，音乐中 C 大调的意思就是基础音调为 C，调式为大调。那么什么是大调、什么是小调呢？

我们刚才说的两个八度音之间基频频率是 2^1 的关系，而一个八度里有 12 个半音的音程，这 12 个半音是按照比值为 $2^{\frac{1}{12}}$ 的等比数列排布的。音程是指两个音之间的间隔。音程的单位是半音或者全音，2 个半音我们就叫 1 个全音，由此推理可得，高一个全音就需要把基频频率乘以 $2^{\frac{2}{12}}$。

我们常说的大、小调又叫自然大、小调，都属于 7 律调式。7 律调式说的就是这 12 个半音里我们只使用其中的 7 个。大调中，每两个音之间的音程大小，依序为全 - 全 - 半 - 全 - 全 - 全 - 半，这就是大调的组成规则，而小调中每两个音之间的音程大小，依序为全 - 半 - 全 - 全 - 半 - 全 - 全。

有了调式我们还需要再指定一个基础音调，比如 C，代表这个音程规律是从哪里开始计数的，这样就可以把调式中所有的音都固定下来了。这个基础音调也叫起始音调，也就是音程间隔开始计算的那个音调，这个音调可以在一个八度里的任意一个半音位置作为开始，也就是说可以有 12 种不同的大调和小调。

你是不是还有点晕？没关系，让我们回过头来再看看表 1。表 1 其实就是 C 大调中的一个八度，基础音调是 C，那么按照大调的“心法口诀”：“全全半，全全全半”那么 C4（do）和 D4（re）之间相差一个全音，频率差其实就是

$$261.6 * 2^{\frac{2}{12}} - 261.6 = 32.1,$$

而 E4（mi）和 F4（fa）之间相差一个半音，频率差是

$$329.6 * 2^{\frac{1}{12}} - 329.6 = 19.6。$$

现在的流行歌主要以自然大调和自然小调为主。但调式除了自然大调和自然小调外还有很多。 比如，中国传统的“宫、商、角、徵、羽”就是一个 5 律调式，分别对应自然大调中的 do、re、mi、so、la 等等，这里就不一一赘述了。

音乐场景问题案例分析

好的，有了前面的音乐信号中的基础知识，我们下面用 2 个案例来看看如何把之前学到的知识用到音乐场景的问题分析中来。

案例 1：客户投诉音乐教学直播时低音鼓一直听不到声音

没有音乐知识的研发同事听到这种投诉，可能会把这个问题当成一个无声问题，从而觉得是不是软件集成哪里出了错误。而我们已经有了前面的音乐知识，看到投诉里提到了低音鼓，就会想到低音鼓是打击乐器，鼓这种乐器一个鼓只有一个音调。而且低音鼓只会发出一个音调的音，且是频率比较低的声音。那么问题就变成了频率比较低的时候听不到声音。

这个时候我们就可以有目的性地询问客户，他们的低音鼓是哪个音调的？假设客户说是 A1 音调的低音鼓。那么我们进一步分析出 A1 比 A4 低 3 个八度。也就是说，A1 的基础频率是 A4 的基频频率除以 8，也就是 55Hz。然后，我们可以先从采集设备能力问一下客户，他们用的是什么设备，这里我们假设客户回答是用的 iPad。接下来，我们可以查一下或者自测一下 iPad 的采集能力。

实际上苹果公司的 iPad 和 iPhone 的最低支持的录音频率是 100Hz，也就是说，设备不支持 55Hz 低频声音的采集。那么这时候我们就可以得出结论，即 iPad 不支持低音鼓这么低频率声音的采集。因此，我们就建议客户更换一个更专业一些的录音设备，比如外接一个可以采集低频的麦克风，来解决问题。

可以看到在这个案例的解决中，我们通过更换采集设备解决了问题，而且里面低音鼓的音调理解起到了关键作用，并帮助我们定位问题，而不是在软件集成等方向上浪费时间。

通过这个问题我们还可以举一反三，解决一系列**乐器基频与采集设备、编解码器和播放设备能力不匹配导致的问题**。例如，乐器是有高频泛音的乐器，比如小提琴，而音频的编 / 解码器只能支持 16kHz 的采样率，这样高频就缺失了，在远端就会感知到部分高音的缺失，这时就需要更换一个支持采样率更高的音频编 / 解码器来解决问题。

案例 2：伴奏升调或者降调

有一款 K 歌的 App，有客户询问有没有办法调节一下伴奏，让唱不上高音的人能唱上去，或者让女生唱男生的歌。这类需求在我们线下 KTV 唱歌的时候也可能会遇到。歌曲在创作的时候伴奏和歌唱的调式都是固定的，有的歌曲高音部分普通人可能唱不到那么高，或者男生的歌的低音部分女生可能唱不了那么低。无论人声唱高了或者低了，伴奏的音调没变，就会出现人声和伴奏的音调不匹配的问题。听起来就会是不和谐的，或者说很容易被听出来“唱走调”。

这个时候我们就可以利用调式的概念。无论大、小调都有一个固定的音程距离，但起始音阶是可以不同的。那么如果我们把伴奏都下调一些，比如都乘以 $2^{\frac{-1}{12}}$ 降一个半音，这样不改变音程距离，但实现了降调，并且这样伴奏还是相同的调式，但本来唱不上去的歌就可以也低一个半音，从而就可以和伴奏匹配了。同理，女生唱男生的歌，就可以把伴奏的频率整体提高一些。比如升一个半音，这样女生就可以用比较高的音调唱男生的歌了。

有了调式的知识，我们就知道调音只有以半音为单位来调节伴奏，才能匹配到正确的调式所需的基频频率。这里还有一个问题要注意，歌曲是有指定的速度和节拍的，如果一个人唱的快了或者慢了，甚至唱多了或漏了几个音，听上去也会感觉是“走调了”，这时改变伴奏的音调并不能解决这种速度或者缺漏音导致的不和谐情况。

我们刚才提到的变调算法有很多，比如 WSOLA，Phase Vocoder 等，是音频音效中常见的算法，在后续的课程中我会和你详细聊一下常见的音效算法，这里就不展开了。

小结

好了，我们来总结一下这节课的内容。音乐中乐器主要包括打击乐器、弦乐器和管乐器，它们的发声原理各不相同。不同的乐器的发音频率、范围也不一样。音乐信号经常会是多个乐器和人声的多信号的混合信号，所以在音乐信号分析时与语音信号分析有所不同。我们通常**需要一个采样率覆盖范围比较广的采集设备和一个采样率支持范围比较广的音频编 / 解码器，来保证音乐信号的频段不会丢失。**

音乐的速度、节拍和调式等都是在音乐创作时就确定了的，任何一个乐器或者人声没有按照指定的基频频率和出现时间就会导致音乐的不和谐，比如“走调”，从而影响听感。

在音乐场景中需要注意的问题其实还有不少是在降噪、回声消除、编 / 解码等音频环节出现的，比如降噪算法可能会导致有些音乐信号的失真甚至消失，而回声消除可能会导致音乐的不流畅。这些我会在今后讲音频处理模块的时候给你详细解读。

思考题

现在我有一个思考题留给你。

人类可以模仿一些乐器的声音，那么从发音原理上来说，有哪些乐器是人类不太可能模仿出来的呢？你可以给 B-box 的艺术家出出难题。

欢迎你在留言区和我分享你的思考和疑惑，你也可以把今天所学分享给身边的朋友，邀请他加入探讨，共同进步。下节课再见。

分享给需要的人，Ta订阅后你可得 20 元现金奖励

生成海报并分享

赞 1　提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇　02 | 如何量化分析语音信号？

下一篇　04 | 如何评价音频质量的好与坏？

精选留言 (5)

写留言

Geek_c32e72
2021-11-29

这个公式什么意思，开2/12根？

展开

作者回复: 这个是2/12、1/12是指数的意思，代表2的多少方

pencilCool
2021-11-28

librosa 已经玩起来了 ，等着老师 下一集

展开

科富
2021-11-26

低音鼓这个案例特别好，如果对频率范围不清楚，从软件上去分析就丢沟了去了

管弦
2021-11-26

某歌手的电音效果，应该是人声模拟不出来的。这种电音效果应该是算法合成的吧。

作者回复: 哈哈，电音效果可以让人发出如机器般精准的音调，确实正常人很难做到。

共 2 条评论 >

springXu
2021-11-26

比起上节课，乐理部分似乎少了数学分析的举例，直觉的多一些。用直觉回答问题，貌似鼓声和小提琴人声难模仿。

展开

作者回复: 直觉对了一半，鼓声是B-box常见的发音哦，而小提琴比较难模仿因为音调比较高。